Working Paper No. 159

GENERALIZED BENDERS DECOMPOSITION

by

ARTHUR M. GEOFFRION

April, 1970

WESTERN MANAGEMENT SCIENCE INSTITUTE

University of California, Los Angeles

38

Statement ONE per guidelines & Kreuson.
A H
22 July 70

GENERALIZED BENDERS DECOMPOSITION

by

Arthur M. Geoffrion
University of California, Los Angeles

April, 1970

Western Management Science Institute

Working Paper No. 159

Prepared for the Nonlinear Programming Symposium at the University of Wisconsin sponsored by the Mathematics Research Center, U.S. Army, May 4-6, 1970.

This research was supported by the National Science Foundation under Grant GP-8740.

Generalized Benders Decomposition

A. M. Geoffrion

University of California, Los Angeles

Abstract

J. F. Benders devised a clever approach for exploiting the structure of mathematical programming problems with "complicating variables". Such problems have the property that temporarily fixing the values of certain variables renders the remaining optimization problem considerably more tractable--perhaps breaking apart into a number of independent smaller problems, or being solvable by a highly efficient known algorithm, or reducing to a convex program in continuous variables whereas the original problem is partly nonconvex or discrete. For the class of problems specifically considered by Benders, fixing the values of the complicating variables reduces the given problem to an ordinary linear program--parameterized, of course, by the value of the complicating variables vector. The algorithm he proposed for finding the optimal value of this vector employs a cutting-plane approach for building up adequate representations of (i) the extremal value of the linear program as a function of the parameterizing vector, and (ii) the set of values of the parameterizing vector for which the linear program is feasible. Linear programming duality theory is employed to derive the natural families of "cuts" characterizing these representations, and the parameterized linear program itself is used to generate what are usually "deepest" cuts for building up the representations.

In this paper, Benders' approach is generalized to a broader class of programs in which the parameterized subproblem need no longer be a linear program. Nonlinear convex duality theory is employed to derive the natural families of "cuts" corresponding to those in Benders' case. The conditions under which such a generalization is possible are examined in detail from both the theoretical and computational viewpoints. An illustrative specialization is made to the variable factor programming problem introduced by R. Wilson, where it appears to offer an especially attractive approach. A computational study is currently under way to evaluate the effectiveness of the generalized Benders decomposition algorithm for several classes of problems.

1. INTRODUCTION

This paper is devoted to problems of the form

$$(1) \qquad \underset{x,y}{\text{Maximize}}\ f(x,y) \ \text{subject to} \ \begin{array}{l} G(x,y) \geq 0 \\ x \in X \\ y \in Y \ , \end{array}$$

where y is a vector of "complicating" variables in the sense that (1) is a much easier optimization problem in x when y is temporarily held fixed. G is an m-vector of constraint functions defined on $X \times Y \subseteq R^{n_1} \times R^{n_2}$.

We have in mind particularly situations such as the following:

(a) for fixed y, (1) separates into a number of independent optimization problems, each involving a different subvector of x

(b) for fixed y, (1) assumes a well-known special structure (such as the classical transportation form) for which efficient solution procedures are available

(c) (1) is not a concave program in x and y jointly, but fixing y renders it so in x.

Such situations abound in practical applications of mathematical programming and in the literature of large-scale optimization, where the central objective is to exploit special structure such as this in order to design effective solution procedures.

It is evident that there are substantial opportunities for achieving computational economies by somehow looking at (1) in y-space rather than in xy-space. We expect that in situation (a), the computations can be

largely decentralized and done in parallel for each of the smaller independent subproblems; in (b), use can be made of available efficient special-purpose algorithms; and in (c), the non-convexities can be treated separately from the convex portion of the problem.

The key idea that enables (1) to be viewed as a problem in y-space is the concept of "projection" [6], sometimes also known as "partitioning." The projection of (1) onto y is:

(2) $$\underset{y}{\text{Maximize}}\ v(y) \quad \text{subject to} \quad y \in Y \cap V ,$$

where

(3) $$v(y) \equiv \underset{x}{\text{Supremum}}\ f(x,y) \quad \text{subject to} \quad \begin{array}{c} G(x,y) \geqq 0 \\ x \in X \end{array}$$

and

(4) $$V \equiv \{y : G(x,y) \geqq 0 \ \text{ for some } \ x \in X\} .$$

Note that $v(y)$ is the optimal value of (1) for fixed y and that, by our designation of y as "complicating" variables, evaluating $v(y)$ is much easier than solving (1) itself. Because we must refer to it so often in the sequel, we introduce the special label $(1y)$ to refer to the optimization problem in (3):

(1y) $$\underset{x \in X}{\text{Maximize}}\ f(x,y) \quad \text{subject to} \quad G(x,y) \geqq 0 .$$

The set V consists of those values of y for which $(1y)$ is feasible; $Y \cap V$ can be thought of as the projection of the feasible region of (1) onto y-space.

It is intuitively clear that the projected problem (2) is equivalent for our purposes to (1). This will be made precise in Theorem 1 below. For now, it is enough to keep in mind that an optimal solution y^* of (2) readily yields an optimal solution (x^*,y^*) of (1), where x^* is any optimizing x in $(1y^*)$.

Benders [1] was one of the first to appreciate the importance of (2) as a route to solving (1). The difficulty with (2), however, is that the function v and the set V are only known implicitly via their definitions (3) and (4). Benders coped with this difficulty by devising a cutting-plane method that builds up a tangential approximation to v and V. His development treats the case in which (1) is a "semilinear" program:

(5a) $$X \equiv \{x : x \geqq 0\} ,$$

(5b) $$f(x,y) \equiv c^t x + \phi(y) ,$$

(5c) $$G(x,y) \equiv Ax + g(y) - b ,$$

where ϕ is a scalar-valued and g a vector-valued function. Under these assumptions, both v and V turn out to have exact representations using only a finite number of tangential approximants. Linear programming duality theory yields a constructive proof of this result based on the fact that (1) is a linear program in x for each fixed y. The rationale of the computational procedure is then evident. See the original paper or [6, Sec. 4.1] for details.

The main result of this paper is an extension of Benders' approach to a more general class of problems. We are able to considerably weaken his semilinearity assumption (5) so as to encompass additional problems of practical interest. This extension is made possible by recent developments in nonlinear duality theory [7]. These results are presented in the following section. In Sec. 3, specialization is made to the Variable Factor Programming Problem previously treated by Wilson [12] by another approach. In addition to its intrinsic interest, this application illustrates an instance in which (1) is not a concave program and yet can be solved by concave programming techniques within the framework of the present approach. Section 4 discusses the assumptions of Sec. 2 and some alternative ways of handling the set V so as to achieve an algorithm that is more primal in nature than might otherwise be the case.

For the reader's convenience, pertinent but not yet widely known results from nonlinear duality theory are summarized in an Appendix.

The notation employed is standard. All vectors are columnar unless transposed.

2. GENERALIZED BENDERS DECOMPOSITION

The generalization of Benders' approach to a class of problems of the form (1) is divided, for pedagogical reasons, into three subsections. The first establishes the "master" problem and its equivalence to (1). The central idea here is to invoke the natural outer tangential representations of v and V after passing to (2). The second subsection demonstrates how the master problem can be solved via a series of "subproblems" which generate the tangential approximants of v and V as needed. Roughly speaking, this is accomplished by obtaining the optimal multiplier vectors for (1y) corresponding to various trial values of y. The Generalized Benders Decomposition Procedure for (1) is then stated and discussed.

2.1 Derivation of the Master Problem

The desired master problem is obtained from (1) by a sequence of three manipulations:

(i) project (1) onto y, resulting in (2),

(ii) invoke the natural dual representation of V in terms of the intersection of a collection of regions that contain it,

(iii) invoke the natural dual representation of v in terms of the pointwise infimum of a collection of functions that dominate it.

Manipulation (i) was already discussed in Sec. 1. The following easy theorem (cf. [6, Sec. 2.1]) shows that (1) and (2) are equivalent for our purposes (note that no assumptions on (1) whatever are needed).

Theorem 1 (Projection). Problem (1) is infeasible or has unbounded optimal value if and only if the same is true of (2). If (x^*,y^*) is optimal in (1), then y^* must be optimal in (2). If y^* is optimal in (2) and x^* achieves the supremum in (3) with $y = y^*$, then (x^*,y^*) is optimal in (1). If $\bar{y}$ is ϵ_1-optimal in (2) and $\bar{x}$ is ϵ_2-optimal in $(1\bar{y})$, then $(\bar{x},\bar{y})$ is $(\epsilon_1+\epsilon_2)$-optimal in (1).

Manipulation (ii) is based on

Theorem 2 (V Representation). Assume that X is a nonempty convex set and that G is concave on X for each fixed $y \in Y$. Assume further that the set $Z_y \equiv \{z \in R^m : G(x,y) \geqq z \text{ for some } x \in X\}$ is closed for each fixed $y \in Y$. Then a point $\bar{y} \in Y$ is also in the set V if and only if $\bar{y}$ satisfies the (infinite) system

$$\left[\underset{x \in X}{\text{Supremum}}\ \lambda^t G(x,y)\right] \geq 0\ , \quad \text{all } \lambda \in \Lambda\ , \tag{6}$$

where $\Lambda \equiv \{\lambda \in R^m : \lambda \geqq 0 \text{ and } \sum_{i=1}^{m} \lambda_i = 1\}$.

Proof. Let $\bar{y}$ be an arbitrary point in Y. It is trivial to verify directly that $\bar{y}$ satisfies (6) if it is in V. The converse can be demonstrated with the help of nonlinear duality theory as follows.

Suppose that $\bar{y}$ satisfies (6). Then

$$\underset{\lambda \in \Lambda}{\text{Infimum}} \left[\underset{x \in X}{\text{Supremum}}\ \lambda^t G(x,\bar{y})\right] \geq 0\ .$$

It follows that

$$\underset{\lambda \geqq 0}{\text{Infimum}} \left[\underset{x \in X}{\text{Supremum}}\ \lambda^t G(x,\bar{y})\right] = 0\ , \tag{7}$$

since the scaling of λ doesn't influence the sign of the bracketed expression and $\lambda = 0$ is allowed in (7). Now (7) simply asserts that the dual with respect to the G constraints of the concave program

$$\text{(8)} \qquad \underset{x \in X}{\text{Maximize}}\ 0^t x \quad \text{subject to} \quad G(x,\bar{y}) \geqq 0$$

has optimal value 0. Recalling that $Z_{\bar{y}}$ is closed, we therefore have from Theorem A.1 of the Appendix that (8) must be feasible and hence that $\bar{y} \in V$.

Q.E.D.

The assumption that Z_y is closed for each $y \in Y$ is not a stringent one. Mild sufficient conditions under which it must hold are given in Sec. 4.1.

Manipulation (iii) is based on

<u>Theorem 3 (v Representation)</u>. Assume that X is a nonempty convex set and that f and G are concave on X for each fixed $y \in Y$. Assume further that, for each fixed $\bar{y} \in Y \cap V$, at least one of the following three conditions holds:

(a) $v(\bar{y})$ is finite and $(1\bar{y})$ possesses an optimal multiplier vector;[1]

(b) $v(\bar{y})$ is finite, $G(x,\bar{y})$ and $f(x,\bar{y})$ are continuous on X, X is closed, and the ϵ-optimal solution set of $(1\bar{y})$ is nonempty and bounded for some $\epsilon \geq 0$;

(c) $v(\bar{y}) = +\infty$.

[1] See the Appendix for the definition of an optimal multiplier vector. Actually, it is enough to consider <u>generalized</u> optimal multiplier vectors (also defined in the Appendix) in order to avoid the implicit assumption that $(1\bar{y})$ must have an optimal solution.

Then the optimal value of (1y) equals that of its dual on $Y \cap V$, that is,

$$\text{(9)} \qquad v(y) = \underset{u \geqq 0}{\text{Infimum}} \left[\underset{x \in X}{\text{Supremum}}\ f(x,y) + u^t G(x,y) \right], \text{ all } y \in Y \cap V .$$

The proof is a direct application of Theorems A.2, A.3, and weak duality. Alternative assumption (a) is one that will very often hold. Many different sufficient conditions, usually called "constraint qualifications," are known which imply it--most of them fairly weak assumptions to preclude pathological cases. Nor is alternative (b) particularly stringent. Of course, boundedness of X or of the feasible region is enough to guarantee that the ϵ-optimal solution set is bounded for any $\epsilon > 0$, and the existence of a unique optimal solution does it for $\epsilon = 0$.

Under the assumptions of Theorems 2 and 3, then, manipulations (i)-(iii) applied to (1) yield the equivalent master problem

$$\underset{y \in Y}{\text{Maximize}} \left[\underset{u \geqq 0}{\text{Infimum}} \left[\underset{x \in X}{\text{Supremum}}\ f(x,y) + u^t G(x,y) \right] \right]$$

subject to (6)

or, using the definition of infimum as the greatest lower bound,

$$\text{(10)} \qquad \underset{\substack{y \in Y \\ y_0}}{\text{Maximize}}\ y_0$$

subject to

$$\text{(10a)} \qquad y_0 \leq \underset{x \in X}{\text{Supremum}} \{f(x,y) + u^t G(x,y)\}, \text{ all } u \geqq 0$$

$$\text{(10b)} \qquad \underset{x \in X}{\text{Supremum}} \{\lambda^t G(x,y)\} \geq 0, \text{ all } \lambda \in \Lambda .$$

2.2 Solving the Master Problem

The most natural strategy for solving the master problem (10), since it has a very large number of constraints, is relaxation [6]: begin by solving a relaxed version of (10) that ignores all but a few of the constraints (10a) and (10b); if the resulting solution does not satisfy all of the ignored constraints, then generate and add to the relaxed problem one or more violated constraints and solve it again; continue in this fashion until a relaxed problem solution satisfies all of the ignored constraints (at which point an optimal solution of (10) has been found), or until a termination criterion signals that a solution of acceptable accuracy has been obtained. Details regarding the termination criterion will be supplied later. The deeper concern is the crucial issue of how a solution to a relaxed version of (10) can be tested for feasibility with respect to the ignored constraints and, in case of infeasibility, how a violated constraint can be generated.

Suppose that $(\hat{y},\hat{y}_0)$ is optimal in a relaxed version of (10). How can this point be tested for feasibility in (10a) and (10b)? From Theorem 2 and the definition of V, we see that $\hat{y}$ satisfies (10b) if and only if $(1\hat{y})$ has a feasible solution. And if $(1\hat{y})$ turns out to be feasible, Theorem 3 implies that $(\hat{y},\hat{y}_0)$ satisfies (10a) if and only if $\hat{y}_0 \leq v(\hat{y})$. Thus $(1\hat{y})$ is the natural subproblem for testing $(\hat{y},\hat{y}_0)$ for feasibility in the master problem. This is in perfect accord with our interest in applications where y is "complicating" in the sense that $(1y)$ is much easier than (1) itself.

Not only is $(1\hat{y})$ the appropriate subproblem for testing the feasibility of $(\hat{y},\hat{y}_0)$ in (10), but almost any reasonable algorithm for $(1\hat{y})$ will yield an index of a violated constraint in the event that $(\hat{y},\hat{y}_0)$ is infeasible. By an "index" of a violated constraint we mean a vector $\hat{u} \geqq 0$ such that

(11a) $$\hat{y}_0 > \underset{x \in X}{\text{Supremum}} \{f(x,\hat{y}) + \hat{u}^t G(x,\hat{y})\}$$

if (10a) is violated, or a vector $\hat{\lambda} \in \Lambda$ such that

(11b) $$\underset{x \in X}{\text{Supremum}} \{\hat{\lambda}^t G(x,\hat{y})\} < 0$$

if (10b) is violated. If $(1\hat{y})$ is infeasible, it can be shown that most dual-type algorithms addressed to it yield such a $\hat{\lambda}$, as do most primal algorithms fitted with a "Phase One" procedure for finding an initial feasible solution if one exists [$\hat{\lambda}$ can be viewed as a convex combination of constraints that has no solution in X]. If $(1\hat{y})$ is feasible and has a finite optimal value, it follows from Theorem A.2 that an optimal multiplier vector satisfies (11a) if one exists and $\hat{y}_0 > v(\hat{y})$. Virtually all modern algorithms applicable to $(1\hat{y})$ produce an optimal multiplier vector as a by-product if one exists, as is usually the case. Nonexistence must be associated either with an unbounded optimal value, in which case one may terminate since the same must then be true of (1), or with a finite optimal value but a pathological condition in which, by (9), $\hat{u}$ nevertheless satisfies (11a) if it comes close enough to

being optimal in the dual of $(1\hat{y})$. Such a $\hat{u}$ will be referred to subsequently as a "near optimal" multiplier vector.

In light of this discussion, it is reasonable to presume henceforth that $(1\hat{y})$ will be addressed with an algorithm that is dual-adequate in the sense that it yields: a vector $\hat{\lambda} \in \Lambda$ satisfying (11b) if $(1\hat{y})$ is infeasible; an optimal multiplier vector $\hat{u}$ if one exists; or a near optimal multiplier vector $\hat{u}$ satisfying (11a) if no optimal multiplier vector exists but $\hat{y}_0$ exceeds the optimal value of $(1\hat{y})$.

Thus far we have shown how $(1\hat{y})$ can be used to test any point $(\hat{y},\hat{y}_0)$ for feasibility in the master problem (10), and to generate an index ($\hat{\lambda}$ or $\hat{u}$) of a violated constraint in the event of infeasibility. There is still a potential difficulty, however, in carrying out the relaxation strategy as a means of solving (10). Namely, the generated constraints may be difficult to obtain in a form suitable for computational purposes even though their indices are known, since they each involve the optimal value of an optimization problem parameterized by y. To overcome this difficulty, it appears necessary to assume that (1) has the following property which, strong as it appears to be, is still much weaker than Benders' semilinearity assumption.

Property P. For every $u \geqq 0$, the supremum of $f(x,y) + u^tG(x,y)$ over X can be taken essentially independently of y, so that there exists a function $L^*(\cdot;u)$ on Y satisfying

$$L^*(y;u) = \underset{x \in X}{\text{Supremum}}\{f(x,y) + u^tG(x,y)\}, \quad y \in Y \tag{12a}$$

which can be obtained explicitly with little or no more effort than is required to evaluate it for a single value of y.

Similarly, for every $\lambda \in \Lambda$, the supremum of $\lambda^t G(x,y)$ over X can be taken essentially independently of y, so that there exists a function $L_*(\cdot;\lambda)$ on Y satisfying

$$(12b) \qquad L_*(y;\lambda) = \underset{x \in X}{\text{Supremum}}\{\lambda^t G(x,y)\},\ y \in Y$$

which can be obtained explicitly with little or no more effort than is required to evaluate it for a single value of y.

One important case in which Property P holds is when f and G are linearly separable in x and y:

$$(13) \qquad \begin{aligned} f(x,y) &\equiv f_1(x) + f_2(y) \ , \\ G(x,y) &\equiv G_1(x) + G_2(y) \ . \end{aligned}$$

Then for any $u \geqq 0$ and $\lambda \in \Lambda$,

$$(14a) \qquad L^*(y;u) = \underset{x \in X}{\text{Supremum}}\{f_1(x)+u^t G_1(x)\} + f_2(y)+u^t G_2(y),\ y \in Y$$

$$(14b) \qquad L_*(y;\lambda) = \underset{x \in X}{\text{Supremum}}\{\lambda^t G_1(x)\} + \lambda^t G_2(y),\ y \in Y\ .$$

The Variable Factor Programming Problem discussed in Sec. 3 shows that Property P can hold even though F and G are not linearly separable in x and y. See Sec. 4.1 for further discussion.

A dual-adequate algorithm for $(1\hat{y})$ that produces $L^*(\cdot;\hat{u})$ or $L_*(\cdot;\hat{\lambda})$ essentially as a by-product as well as $\hat{u}$ or $\hat{\lambda}$ will be

called L/dual-adequate. In the presence of Property P, L/dual-adequacy usually requires very little if anything more of an algorithm for $(1\hat{y})$ than dual-adequacy does. When (13) holds, for example, by (9) and (14a) we have

$$L^*(y;\hat{u}) = v(\hat{y})-f_2(\hat{y}) - \hat{u}^t G_2(\hat{y}) + f_2(y)+\hat{u}^t G_2(y),\ y \in Y$$

when $\hat{u}$ is an optimal multiplier vector [for then $\hat{u}$ achieves the infimum in (9)]. This equation is only approximate when an optimal multiplier vector does not exist, but the approximation can be made arbitrarily accurate by making $\hat{u}$ nearly enough optimal in the dual of $(1\hat{y})$ [cf. Th. 3]. Thus $L^*(y;\hat{u})$ is obtained without even having to explicitly perform the supremum in (14a). A similarly satisfactory situation usually exists regarding $L_*(y;\hat{\lambda})$, although a detailed justification of this statement would depend upon whether the algorithm produces $\hat{\lambda}$ by a Phase One or dual-type approach when $(1\hat{y})$ is infeasible. In the context of Variable Factor Programming, it will be seen in Sec. 3 that any algorithm for $(1\hat{y})$ is L/dual-adequate so long as it produces an optimal solution $\hat{x}$ and multiplier vector $\hat{u}$ (both of which necessarily exist for any $\hat{y}$).

We can now formally state the relaxation procedure for solving the master problem (10). It is assumed henceforth that the hypotheses of Theorems 2 and 3 hold, that Property P holds, and that an L/dual-adequate algorithm for $(1y)$ is available.

Generalized Benders Decomposition Procedure

Step 1. Let a point $\bar{y}$ in $Y \cap V$ be known. Solve the subproblem (1$\bar{y}$) by any L/dual-adequate algorithm (terminate if the optimal value is $+\infty$) to obtain an optimal (or near optimal) multiplier vector $\bar{u}$ and the function $L^*(y;\bar{u})$. Put $p = 1$, $q = 0$, $u^1 = \bar{u}$, $LBD = v(\bar{y})$. Select the convergence tolerance parameter $\epsilon > 0$.

Step 2. Solve the current relaxed master problem,

$$\text{(15)} \qquad \underset{\substack{y \in Y \\ y_0}}{\text{Maximize}} \; y_0 \text{ subject to } y_0 \le L^*(y;u^j),\ j = 1,\dots,p$$
$$L_*(y;\lambda^j) \ge 0,\ j = 1,\dots,q\ ,$$

by any applicable algorithm. Let $(\hat{y},\hat{y}_0)$ be an optimal solution. If $LBD \ge \hat{y}_0 - \epsilon$, terminate.

Step 3. Restart the L/dual-adequate algorithm for the revised subproblem (1$\hat{y}$). One of the following three cases must occur.

A. $v(\hat{y}) = +\infty$. Terminate.

B. $v(\hat{y})$ is finite. If $v(\hat{y}) \ge \hat{y}_0 - \epsilon$, terminate. Otherwise, determine an optimal multiplier vector $\hat{u}$ (if none exists, a near optimal multiplier vector satisfying (11a) will do) and the function $L^*(y;\hat{u})$. Increase p by 1 and put $u^p = \hat{u}$. If $v(\hat{y}) > LBD$, put $LBD = v(\hat{y})$. Erase $\hat{y}$ and $\hat{y}_0$ and return to Step 2.

C. (1$\hat{y}$) is infeasible. Determine $\hat{\lambda}$ in Λ satisfying (11b), and the function $L_*(y;\hat{\lambda})$. Increase q by 1 and put $\lambda^q = \hat{\lambda}$. Erase $\hat{y}$ and $\hat{y}_0$ and return to Step 2.

A few remarks are in order.

Remark 1. It is implicitly assumed in the statement of Step 2 that an optimal solution of (15) exists. This problem is necessarily feasible in view of the assumed existence of the point $\bar{y}$ in Step 1, but to preclude an unbounded optimum during the early executions of Step 2 it may be necessary to temporarily bound y or y_0 artificially.

Remark 2. Step 1 is set up to begin at a known point $\bar{y}$ in $Y \cap V$ because this is believed to be the situation most likely to be encountered in applications. There is an advantage to utilizing experience and mathematical or physical insight in order to provide a good initial feasible solution. If such a point is unknown, however, Step 1 could be altered in the obvious way to accomodate an infeasible initial point (cf. Step 3C). An alternative would be to make the procedure as it stands function in a "Phase One" mode in order to find a point $\bar{y}$ in $Y \cap V$.

Remark 3. The termination conditions can be understood as follows, in view of Theorem 1. Termination at Step 1 means that there is a sequence $< x^{\nu} >$ feasible in $(1\bar{y})$ such that $< f(x^{\nu},\bar{y}) > \to +\infty$. Thus the sequence $< (x^{\nu},\bar{y}) >$ is feasible in (1) and drives its value to $+\infty$. Termination at Step 3A is of the same type. Termination at Step 3B implies that $\hat{y}$ is ϵ-optimal in (2), for $\hat{y}_0$ is obviously an upper bound on the optimal value of (2) [it is the optimal value of the equivalent master problem (10) with some of its constraints deleted]. Thus any optimal solution $\hat{x}$ of $(1\hat{y})$ yields an ϵ-optimal solution $(\hat{x},\hat{y})$ of (1) [if only a δ-optimal solution of $(1\hat{y})$ can be computed, the result is $(\epsilon+\delta)$-optimal

in (1)]. Termination at Step 2 is similar, except that LBD plays the role of $v(\hat{y})$; LBD is always set at the greatest optimal value of (1y) found at any previous execution of Step 1 or Step 3B, and so is the best known <u>lower bound</u> on the optimal value of (2). When LBD $\geq \hat{y}_0 - \epsilon$, the subproblem corresponding to LBD yields an ϵ-optimal solution of (1). Note that while the sequence of values for $\hat{y}_0$ found at successive executions of Step 2 must be monotone nonincreasing, the sequence of values for $v(\hat{y})$ found at Step 3B need <u>not</u> be monotone nondecreasing. This is the reason for introducing LBD.

<u>Remark 4</u>. In practice, a reasonably efficient algorithm must be available for solving the relaxed master problem (15) at Step 2. Except for certain very special cases, this normally requires that the set Y be convex and that the L^* and L_* functions be concave on Y (sufficient conditions are given in Sec. 4.1). Then (15) becomes an ordinary concave programming problem to which a number of available algorithms apply. It should be mentioned, however, that the case where Y is a discrete set is also of some significance. One of the main points of Benders' original paper is that his approach permits solution of mixed integer programs when y is identified with the integer variables and x is identified with the continuous variables--provided, of course, that a (nearly) pure integer programming procedure is available to solve (15).

<u>Remark 5</u>. For computational efficiency it is important that the algorithms used by the procedure to solve (1y) and (15) be able to restart each time taking advantage of the most recent previous solution.

Remark 6. In practice, the number of constraints in (15) probably needn't be allowed to build up indefinitely. The important recent work of Eaves and Zangwill [3] and Topkis [10, 11] suggests that amply satisfied constraints can usually be dropped after each execution of Step 2. In the special case where (10a) and (10b) are each equivalent to a finite subset of themselves, as in Benders' case (5), a simple argument [6, Sec. 3.3] is enough to justify dropping amply satisfied constraints provided (1) is a concave program.

Remark 7. It is of some interest and comfort to know that the violated constraints generated at Step 3 are usually the most violated (or nearly so) among all violated constraints. When $\hat{u}$ is an optimal multiplier vector at Step 3B, it follows from Theorem A.2 that it indexes a constraint among (10a) that is most violated at $(\hat{y},\hat{y}_0)$. When no optimal multiplier vector for $(1\hat{y})$ exists, how nearly $\hat{u}$ comes to indexing a most violated constraint depends solely on how nearly it solves the dual problem of $(1\hat{y})$. Similarly, how close $\hat{\lambda}$ comes to indexing a most violated constraint among (10b) at Step 3C depends solely on how close it comes to solving the dual of (8) with $\bar{y}$ equal to $\hat{y}$ (with the dual vector normalized to Λ).

Remark 8. A major potential difficulty with the procedure is that Step 3C may occur too frequently by comparison with Step 3B, with the result that the procedure spends most of its effort trying to remain feasible in V to the neglect of the maximization aspect of the problem. In Benders' case (5), it turns out that Step 3C can never occur more than a finite number of times. Although "finite" can be very large, at least this

precludes Step 3C from occuring infinitely often. In the case studied in Sec. 3, V turns out to be the whole space. This means that Step 3C can never occur at all. Thus the frequency with which Step 3C occurs, while it need not be excessive, does have the effect of limiting the scope of problems for which the procedure is effective. In Sec. 4.2 an attempt is made to delimit this scope more sharply, and to point out modifications of the procedure that would help to expand it.

3. APPLICATION TO VARIABLE FACTOR PROGRAMMING

Wilson [12] has defined the Variable Factor Programming Problem as follows:

(16) $$\underset{x,y}{\text{Maximize}} \sum_{i=1}^{n_2} y_i f_i(x^i)$$

subject to

(16a) $$\sum_{i=1}^{n_2} x^i y_i \leqq c$$

(16b) $$x^i \geqq 0, \quad i = 1,\ldots,n_2$$

(16c) $$Ay \leqq b, \quad y \geqq 0 ,$$

where x^i is an m-vector [x is the $n_2 \cdot m$-vector $(x^1,\ldots,x^{n_2})$]. One may interpret y_i as the level of activity i, and f_i as the profit coefficient or yield rate for activity i when x^i is the allocation of variable factors (or stimulants) per unit of activity i. Constraint (16a) requires the total allocation of variable factors to be within an availability vector c.

Obviously (16) is a proper generalization of the general linear programming problem. As pointed out by Wilson, it arises in petroleum refining and blending (tetra-ethyl lead is a stimulant), certain pre-processing industries, and agricultural planning. In the later case, for instance, y_i might represent the number of acres to be planted with crop i, and x^i might represent the amounts of various kinds of fertilizer, irrigation water, tilling labor, etc., to be used on each acre of crop i.

It will be assumed that each f_i is a continuous concave function on the nonnegative orthant, i.e., that the marginal productivity of the variable factors is nonincreasing. It is also assumed without loss of generality that each component of c is positive (if some component of c were 0, the corresponding variable factor could be dropped from the problem). It then follows that the assumptions of Sec. 2 hold. In fact, a major simplification even arises because the set V turns out to be all of R^{n_2}. This is so because $c \geqq 0$ and $x^1 = \ldots = x^{n_2} = 0$ satisfies (16a) for any choice of y. The assumptions underlying the Generalized Benders Decomposition Procedure are easily verified as follows. Theorem 2, of course, is unnecessary. In Theorem 3, the initial assumptions are immediate. Alternative assumption (a) holds for all $\bar{y} \geqq 0$ because the subproblem $(1\bar{y})$, which specializes to

$$(17\bar{y}) \qquad \underset{x^1 \geqq 0, \ldots, x^{n_2} \geqq 0}{\text{Maximize}} \quad \sum_{i=1}^{n_2} \bar{y}_i f_i(x^i)$$

subject to

$$\sum_{i=1}^{n_2} x^i \bar{y}_i \leqq c \; ,$$

is a concave program (remember that $y \geqq 0$) with a finite optimal solution[2] which also satisfies Slater's interiority constraint qualification ($x^1 = \ldots = x^{n_2} = 0$ is interior since $c > 0$). The relevant part of

---

[2] The objective function is continuous, x^i is arbitrary for i such that $\bar{y}_i = 0$, and x^i is bounded by $0 \leqq x^i \leqq c/\bar{y}_i$ for i such that $\bar{y}_i > 0$.

Property P, namely the first part, also holds since for all $y \geqq 0$ we have

$$(18) \qquad \underset{x \in X}{\text{Supremum}} \{f(x,y) + u^tG(x,y)\}$$

$$= \underset{x^1 \geqq 0,\ldots,x^{n_2} \geqq 0}{\text{Supremum}} \left\{ \sum_{i=1}^{n_2} y_i f_i(x^i) + u^t\left(-\sum_{i=1}^{n_2} x^i y_i + c\right)\right\}$$

$$= u^tc + \sum_{i=1}^{n_2} y_i \underset{x^i \geqq 0}{\text{Supremum}} \{f_i(x^i) - u^tx^i\} = L^*(y;u)$$

Note that $L^*(y;u)$ is linear in y. Finally, we note that any algorithm for solving ($17\bar{y}$) at Step 1 or 3 is dual-adequate so long as it produces an optimal solution $\bar{x}$ and an optimal multiplier vector $\bar{u}$ (the existence of both for any $\bar{y}$ has been demonstrated above). Such an algorithm is also L/dual-adequate, because

$$(19) \qquad \underset{x^i \geqq 0}{\text{Supremum}} \{f_i(x^i) - \bar{u}^tx^i\} = f_i(\bar{x}^i) - \bar{u}^t\bar{x}^i$$

holds for all i such that $\bar{y}_i > 0$; for i such that $\bar{y}_i = 0$, the supremum will have to be computed. Thus $L^*(y;\bar{u})$ is at hand directly from (18) and (19) with little if any extra work once the subproblem ($17\bar{y}$) is solved.

It would seem that Generalized Benders Decomposition is particularly well suited to Variable Factor Programs. Steps 3A and 3C cannot occur; the relaxed master program to be solved at Step 2,

$$(20) \qquad \underset{\substack{y \geqq 0 \\ y_0}}{\text{Maximum}} \; y_0 \;\text{ subject to }\; Ay \leqq b$$
$$y_0 \le L^*(y;u^j),\; j = 1,\ldots,p \,,$$

is an ordinary linear program (L^* is linear in y for fixed u); and the subproblem (17y) to be solved at Steps 1 and 3 is a concave program with linear constraints and an objective function that is linearly separable in the x^i's. This subproblem even has a natural and meaningful interpretation in its own right as the problem of determining the optimal use of variable factors given a tentative choice of activity levels. The violated constraints generated are always most violated constraints (see Remark 7).

Only trivial modifications of the above analysis are required to accomodate additional constraints of the form $y \in Y$ and $x^i \in X^i$, so long as Y and each X^i is convex and X^i contains the origin. Of course, (20) may no longer be a linear program because it would then include the constraint $y \in Y$.

4. DISCUSSION

The first subsection discusses an assumption of Theorem 2 and some of the implications of Property P. The second subsection elaborates on Remark 8 of Sec. 2 concerning the possibility that the Generalized Benders Decomposition Procedure may spend too much effort on maintaining feasibility of the subproblem. The final subsection mentions a computational study now under way.

4.1 Discussion of Assumptions of Sec. 2

Theorem 2 assumes that the set

$$Z_y \equiv \{z \in R^m: G(x,y) \geqq z \text{ for some } x \in X\}$$

must be closed for each $y \in Y$. At first glance it might seem that this would be so if G were continuous on X for each fixed y in Y and X were closed, but the following example shows that this is not sufficient: let $m = 1$, $X = [1,\infty)$, and $G(x,y) = -1/x$; then clearly $Z_y = (-\infty, 0)$, which is not closed. If X is bounded as well as closed and G is continuous on X for each $y \in Y$, however, it is easy to verify that Z_y must be closed for all $y \in Y$. It can also be shown (with the help of Lemma 6 of [7]) that these conditions remain sufficient if the boundedness of X is replaced by the condition: X is convex and, for each fixed $y \in Y$, G is concave on X and there exists a point z_y such that the set $\{x \in X : G(x,y) \geqq z_y\}$ is bounded and nonempty. Since X is required to be convex and G is required to be concave on X for fixed $y \in Y$ by Theorem 3 anyway, this condition is a useful weakening

of the requirement that X be bounded. For example, the condition holds if, for each fixed y in Y, at least one of the component functions of G has a bounded set of maximizers over X (in this case, select the corresponding component of z_y to be the maximal value and let the other components be arbitrarily small).

The significance of the requirement of Property P that the extrema in (12a) and (12b) "can be taken essentially independently of y" deserves further elaboration. We have already seen two classes of problems for which this requirement holds, namely linear separability in x and y [see (14)] and Variable Factor Programming [see (18)]. A third class of problems for which it holds are those in which $f(x,y)+u^tG(x,y)$ can be written as $Q(h(x,u),y,u)$ for any $x \in X$, $y \in Y$, and $u \geqq 0$, where Q is increasing in its first component and h is a scalar function of x and u. Then clearly

(21) $$L^*(y;u) = Q(\underset{x \in X}{\text{Supremum}}\{h(x,u)\},y,u), \quad y \in Y .$$

A similar representation would also have to be available for $\lambda^tG(x,y)$ if the second part of P is to hold.

When the achievement of the extrema in (12a) and (12b) is not in question, the following version of Property P is appropriate: for every $u \geqq 0$ there exists a point $x_u \in X$ such that

(22a) $$\underset{x \in X}{\text{Supremum}}\{f(x,y)+u^tG(x,y)\} = f(x_u,y)+u^tG(x_u,y), \quad y \in Y ,$$

and for every $\lambda \in \Lambda$ there exists a point $x_\lambda \in X$ such that

(22b) $$\underset{x \in X}{\text{Supremum}} \{\lambda^t G(x,y)\} = \lambda^t G(x_\lambda, y), \quad y \in Y .$$

Then $L^*(y;u)$ is available directly from x_u, and $L_*(y;\lambda)$ directly from x_λ. Let this version be called Property P'. One of the consequences of P' is that, by weak duality, the subproblem (1y) must have a finite optimal value whenever it is feasible. Thus Step 3A of the Generalized Benders Decomposition Procedure cannot occur, and termination cannot occur at Step 1. It is helpful to keep in mind that the usual situation, but by no means the only possible one, is where $\hat{u}$ is an optimal multiplier vector for $(1\hat{y})$ and $x_{\hat{u}}$ is an optimal solution.

One conjecture concerning Property P that has probably already occurred to the reader is that the first part implies the second part. This seems plausible since for any $\lambda \in \Lambda$ in the second part of P one may take $u = \theta\lambda$ in the first part with θ so large that the influence of f becomes inconsequential by comparison with that of $\theta\lambda^t G$, with the result that $L_*(y;\lambda)$ approximately equals $\frac{1}{\theta} L^*(y;\theta\lambda)$ for large θ. This conjecture can indeed be verified when X is closed and bounded and f and G are continuous on X for each fixed $y \in Y$. The proof is most easily carried out in terms of P'. The boundedness of X can be weakened somewhat along the lines suggested at the end of the first paragraph of this subsection.

Finally, it is appropriate to recite some sufficient conditions for $L^*(L_*)$ to be concave as a function of y for fixed $u(\lambda)$. As noted in Remark 4 in Sec. 2, this is usually necessary if the relaxed master problem (15) is to be solvable by an available algorithm. Suppose

that Y is a convex set. We consider several possible cases. When f and G are linearly separable in x and y, it is evident from (14) that the desired conclusion holds if f_2 and G_2 are concave on Y. In Variable Factor Programming, (18) shows that L^* is even <u>linear</u> on Y. When the Q-representation introduced above holds, L^* is concave if Q is concave in its y-argument, and L_* is concave if the analog of Q is concave in its y-argument. And if the slightly more stringent Property P' holds, (22) reveals that the desired conclusion obtains if f and G are concave on Y for each fixed $x \in X$. With regard to this last condition, it is perhaps unnecessary to say that marginal concavity of f and G in x for fixed y and in y for fixed x does not imply joint concavity on $X \times Y$, although of course the converse does hold. It is useful to know, however, that this stronger joint condition does imply without any further qualification that L^* and L_* are concave on Y. The proof follows the lines of Theorem 2 in [6].

It is interesting to note that under any of these sufficient conditions for the concavity of L^* and L_*, not only is (15) a concave program, but so is the original projected problem (2): by Theorem 2, $V \cap Y$ is the convex set $\{y \in Y : L_*(y;\lambda) \geq 0, \text{ all } \lambda \in \Lambda\}$; and by Theorem 3, $v(y) = \underset{u \geqq 0}{\text{Infimum}}\ L^*(y;u)$ is concave on $Y \cap V$ since it is the infimum of a collection of concave functions. <u>Thus (2) can be a concave program even though (1) is not a concave program</u>. This is certainly true in Variable Factor Programming, as we have already pointed out. It can also be true, for example, when the sufficient condition

associated with Property P' holds.

4.2 On Maintaining Feasibility in the Subproblem

As discussed in Remark 8 of Sec. 2, the effectiveness of the Generalized Benders Decomposition Procedure is diminished if Step 3C arises too often, i.e., if the solutions of the relaxed master problem at Step 2 too often lie outside of V. It was pointed out that this difficulty will not materialize for applications such as Variable Factor Programming in which $Y \subset V$. In other applications it may be possible to specify, based on prior experience or by physical or mathematical insight, a region entirely within V known to contain an optimal solution of (2). This information would be used to redefine Y so that $Y \subset V$ holds. If this cannot be done, sometimes it is still possible to obtain V explicitly before commencing the procedure. This possibility is mainly limited to cases in which G or y is of very low dimensionality. For example, if $m = 1$ (i.e., G is a scalar-valued function) then V can be represented by the single constraint $L_*(y;1) \geq 0$. The case $n_2 = 1$ (i.e., y is a scalar) is also manageable when (2) is a concave program--see the sufficient conditions given in Sec. 4.1--because then V is a convex subset of the real line and must therefore be a simple interval. It would not be difficult to determine the endpoints.

Step 3C will not occur at all in the above cases because there was no need to approximate V. The next best situation is where Step 3C is limited to a finite number of possible occurrences because the representation of Theorem 2 is essentially a finite one. This is so (and he

recognized it) under Benders' assumptions (5a) and (5c):

$$\underset{x \geqq 0}{\text{Supremum}}\ \lambda^t[Ax + g(y) - b]$$

$$= \lambda^t[g(y) - b] + \underset{x \geqq 0}{\text{Supremum}}\ (\lambda^t A)x$$

$$= \lambda^t[g(y) - b] + \begin{cases} +\infty & \text{if } \lambda^t A \nleqq 0 \\ 0 & \text{if } \lambda^t A \leqq 0 \ . \end{cases}$$

It follows that V is represented by the finite system

(23) $$(\lambda^j)^t[g(y) - b] \geq 0, \quad j = 1,\ldots,r \ ,$$

where $\{\lambda^1,\ldots,\lambda^r\}$ is the set of extreme points of the convex polyhedron

(24) $$\{\lambda \geqq 0 : \lambda^t A \leqq 0 \text{ and } \sum_{i=1}^{m} \lambda_i = 1\}$$

(if this polyhedron is empty, then $V = R^{n_2}$). Provided the L/dual-adequate algorithm selected for Step 3 is arranged so as to generate only extreme points of the polyhedron (24), Step 3C can occur no more than r times. Hopefully only a small subset of all r extreme points will ever have to be generated. If this is not the case, however, then it may be wise to modify the procedure so as to essentially absorb Step 3C into Step 2 so that priority is given to assuring that $\hat{y} \in V$ when the relaxed master problem is solved. This can be done if (i) a parametric programming algorithm for the subproblem (1y) is available that can accomodate linear perturbations of y [4, 8, 9], and (ii) a feasible directions algorithm is used for the relaxed master problem (15). Since most

feasible directions algorithms require knowledge only of the constraints binding (or very nearly binding) at the current feasible solution and alter this solution by a sequence of linear moves, it is possible to carry out the feasible directions algorithm for (15) with $q = r$ without having to generate all r extreme points in advance. The relevant extreme points would be generated as needed by the parametric algorithm for (1y), which would track the optimal solution of the subproblem during each linear move of y and produce an appropriate new constraint [extreme point of (24)] whenever (1y) is about to go infeasible (i.e., whenever y is about to leave V). Space limitations preclude giving further details here.

There may be other useful conditions under which V has a finite representation, but in general (10b) could be an infinite system with the result that Step 3C may occur infinitely many times in a row. In this event one may have to abandon the procedure entirely unless a modification such as the following is sufficiently effective in circumventing the difficulty. Only a brief outline can be given here of the suggested modification, which would be to treat the set $Y \cap V$ in (2) by inner rather than outer approximation, assuming that it is convex. In the terminology of [6], the idea would be to solve (2) by outer linearization/relaxation applied to the objective function v only (instead of to both v and $Y \cap V$), with the relaxed master problems solved by inner linearization/restriction applied to the set $Y \cap V$ only.[3] This possibility

[3] Inner linearization/restriction is a conceptual approach due to Dantzig and Wolfe [2, Ch. 24]. See [6, Sec. 4.3] for the extension of this approach needed here.

was examined and seemed quite attractive in [5, Sec. 3.1] for a structure in which G was separable in x and y and linear in y. The advantage is that the new relaxed master problem,

$$\text{(25)} \qquad \underset{\substack{y \in Y \cap V \\ y_0}}{\text{Maximize}} \; y_0 \quad \text{subject to} \quad y_0 \leq L^*(y;u^j), \; j = 1,\dots,p \; ,$$

would be solved primally for $\hat{y}$; thus $\hat{y}$ would necessarily be in V and so $(1\hat{y})$ could not be infeasible.[4/] The price paid for eliminating Step 3C entirely in this way is, of course, more work at Step 2.

4.3 Computational Experience

A computational study is now under way to compare, for several classes of test problems, the efficiency of Generalized Benders Decomposition against a direct nonlinear programming attack. A single sophisticated nonlinear programming algorithm (NONLIN, due to Prof. G. W. Graves at UCLA) is being used both for the direct attack and for the relaxed master problems and the subproblems. The results of this study will be reported at a later date.

[4/] In practice, the inner linearization/restriction algorithm applied to (25) would be terminated suboptimally, especially during the early executions of Step 2 (p small). Since an upper as well as lower bound is available to use in a termination criterion for (25), however, it is a simple matter to increase the required accuracy at successive executions so as to maintain control over the convergence of the entire procedure.

APPENDIX

Summary of Relevant Results from Nonlinear Duality Theory

All results will be stated in terms of the standard primal problem

$$\text{(A.1)} \qquad \underset{x \in X}{\text{Maximize}}\ f(x) \ \text{ subject to } \ g_i(x) \geq 0, \quad i = 1,\ldots,m\ ,$$

where it is assumed that f and each function g_i is concave on the non-empty convex set $X \subseteq R^n$. See [7] for further details.

The dual of (A.1) with respect to the g_i constraints is

$$\text{(A.2)} \qquad \underset{u \geqq 0}{\text{Minimize}} \left[\underset{x \in X}{\text{Supremum}}\ f(x) + \sum_{i=1}^{m} u_i g_i(x) \right],$$

where $u = (u_1,\ldots,u_m)$ is a vector of dual variables. By the Weak Duality Theorem [ibid., Th.2], any feasible solution of the primal must have a value no greater than the value of any feasible solution of the dual.

If $\bar{x}$ is an optimal solution of the primal, an optimal multiplier vector $\bar{u}$ is defined to be any nonnegative vector satisfying: $\sum_{i=1}^{m} \bar{u}_i g_i(\bar{x}) = 0$ and $\bar{x}$ maximizes $f(x) + \sum_{i=1}^{m} \bar{u}_i g_i(x)$ over X. To preclude assuming that the optimal value of the primal is actually achieved by some $\bar{x}$, for many purposes it is enough to work with the concept of a generalized optimal multiplier vector: a nonnegative vector $\bar{u}$ such that, for every scalar $\epsilon > 0$, there exists a point x_ϵ feasible in the primal problem satisfying the two conditions (i) x_ϵ is an ϵ-optimal maximizer of $f(x) + \sum_{i=1}^{m} \bar{u}_i g_i(x)$ over X, and

(ii) $\sum_{i=1}^{m} \bar{u}_i g_i(x_\epsilon) \le \epsilon$. Every optimal multiplier vector is also a generalized optimal multiplier vector.

Theorem A.1. If $\{z \in R^m : g_i(x) \geq z_i,\ i = 1,\ldots,m,$ for some $x \in X\}$ is closed and the optimal value of the dual is finite, then the primal problem must be feasible.

Proof. This is an immediate corollary of Theorem 5 [ibid.].

Theorem A.2. If $\bar{u}$ is a (generalized) optimal multiplier vector for the primal problem, then $\bar{u}$ is also an optimal solution of the dual and the optimal values of primal and dual are equal.

Proof. This follows immediately from Lemmas 3 and 4 [ibid.], and the discussion just before the latter.

It is also true that if the primal has no generalized optimal multiplier vector and yet the optimal values of primal and dual are equal, then the dual can have no optimal solution (by inspection of Diagram 1 [ibid.]).

Theorem A.3. Assume X to be closed, f and each g_i to be continuous on X, the optimal value of (A.1) to be finite and the set

$$\{x \in X : g_i(x) \geq 0,\ i = 1,\ldots,m, \text{ and } f(x) \geq \alpha\}$$

to be bounded and nonempty for some scalar α no greater than the optimal value of (A.1). Then the optimal values of the primal and dual problems are equal.

Proof. This follows immediately from Theorems 7 and 8 [ibid.].

REFERENCES

1. Benders, J. F., "Partitioning Procedures for Solving Mixed-Variables Programming Problems," Numerische Mathematik, 4 (1962), pp. 238-252.

2. Dantzig, G. B., Linear Programming and Extensions, Princeton University Press, Princeton, New Jersey, 1963.

3. Eaves, B. C., and W. I. Zangwill, "Generalized Cutting Plane Algorithms," Working Paper No. 274, Center for Research in Management Science, University of California, Berkeley, July 1969.

4. Geoffrion, A. M., "Strictly Concave Parametric Programming, Parts I and II," Management Science, 13, 3 (November 1966) and 5 (January 1967), pp. 244-253 and 359-370.

5. Geoffrion, A. M., "Primal Resource-Directive Approaches for Optimizing Nonlinear Decomposable Systems," The RAND Corporation, RM-5829-PR, December 1968. To be published in Operations Research.

6. Geoffrion, A. M., "Elements of Large-Scale Mathematical Programming," Working Paper No. 144, Western Management Science Institute, University of California, Los Angeles, September 1969. To be published in Management Science, July 1970.

7. Geoffrion, A. M., "Duality in Nonlinear Programming: A Simplified Applications-Oriented Development," Working Paper No. 150, Western Management Science Institute, University of California, Los Angeles, November 1969. To be published in SIAM Review.

8. Grigoriadis, M. D., and K. Ritter, "A Parametric Method for Semi-Definite Quadratic Programs," J. SIAM Control, 7, 4 (November 1969), pp. 559-577.

9. Silverman, G. J., "Primal Decomposition of Mathematical Programs by Resource Allocation," Technical Memorandum No. 116, Operations Research Department, Case Western Reserve University, Cleveland, August 1968.

10. Topkis, D. M., "Cutting-Plane Methods Without Nested Constraint Sets," ORC 69-14, Operations Research Center, University of California, Berkeley, June 1969. To be published in Operations Research.

11. Topkis, D. M., "A Note on Cutting-Plane Methods Without Nested Constraint Sets," ORC 69-36, Operations Research Center, University of California, Berkeley, December 1969.

12. Wilson, R., "Programming Variable Factors," Management Science, 13, 1 (September 1966), pp. 144-151.